PERGAMON

Vision Research 39 (1999) 19–30

Vision Research

Local motion detectors cannot account for the detectability of an extended trajectory in noise

Preeti Verghese [a,*], Scott N.J. Watamaniuk [b], Suzanne P. McKee [a], Norberto M. Grzywacz [a]

[a] *Smith-Kettlewell Eye Research Institute, 2232 Webster Street, San Francisco, CA 94115-1821, USA*
[b] *Department of Psychology, Wright State University, Dayton, OH 45435, USA*

Received 22 July 1997; received in revised form 1 December 1997

Abstract

Previous work has shown that a single dot moving in a consistent direction is easily detected among noise dots in Brownian motion (Watamaniuk et al., Vis Res 1995;35:65–77). In this study we calculated the predictions of a commonly-used psychophysical motion model for a motion trajectory in noise. This model assumes local motion energy detectors optimally tuned to the signal, followed by a decision stage that implements the maximum rule. We first show that local motion detectors do indeed explain the detectability of brief trajectories (100 ms) that fall within a single unit, but that they severely underestimate the detectability of extended trajectories that span multiple units. For instance, a 200 ms trajectory is approximately three times more detectable than two isolated 100 ms trajectories presented together within an equivalent temporal interval. This result suggests a nonlinear interaction among local motion units. This interaction is not restricted to linear trajectories because circular trajectories with curvatures larger than 1° are almost as detectable as linear trajectories. Our data are consistent with a flexible network that feeds forward excitation among units tuned to similar directions of motion.

Keywords: Brownian motion; Motion energy; Local detector; Trajectory network

1. Introduction

Most of the current knowledge of visual motion processing is about the properties of local motion units, which are well-modeled as linear filters in space and time. The responses of these local motion units must be combined over space and time to guide locomotion and to aid object recognition, but little is known about the neural mechanisms responsible for integrating these local signals. Due to momentum, objects do not usually change speed or direction abruptly. Therefore, one way to maintain object identity is to enhance the signals in local motion units that are compatible with motion in a consistent direction, particularly the signals in sequentially-stimulated units, i.e. a motion trajectory. Previous work has shown that trajectory motion is easily detected amidst Brownian motion noise (Watamaniuk, McKee & Grzywacz, 1995). Here we ask whether this detectability can be predicted by local motion units acting independently or whether it requires a network of local units with facilitatory connections between units tuned to the same direction of motion.

There are several studies that suggest that signals moving in a consistent direction are enhanced. For instance, multi-frame motion is much more detectable than two-frame motion spanning the same distance over time (Nakayama & Silverman, 1984; McKee & Welch, 1985; Snowden & Braddick, 1989). One mechanism that can account for these data is a 'sequential recruitment' of motion signals in the direction of motion. An alternate explanation is that the larger number of samples in the multi-frame case provide more stimulus energy to a local motion unit. Other studies have shown that the direction of ambiguous motion can be resolved by embedding the ambiguous stimuli in a sequence moving in a consistent direction, a phenomenon called 'motion inertia' (Ramachandran & Anstis, 1983; Anstis & Ramachandran, 1987). Motion inertia can be taken as evidence for the accumulation of

* Corresponding Author. Fax: + 1 415 5611610; e-mail: preeti@skivs.ski.org.

0042-6989/98/$ - see front matter
PII: S0042-6989(98)00033-9

a trajectory signal over time that persists despite a brief interjection of stimuli with ambiguous direction. An alternate explanation is that motion inertia is due to a large motion unit summing direction inputs, of which only a subset are ambiguous. However, in the experimental conditions of the Watamaniuk et al. (1995) study cited above, a detector large enough to cover the entire signal trajectory also sees a large number of noise dots and thus has a poor signal-to-noise ratio (Grzywacz, Watamaniuk & McKee, 1995; also see Fig. 1 below). There is other evidence that argues against large detectors mediating trajectory detection. Circular trajectories of curvature larger than 1° are detected almost as easily as linear trajectories (Grzywacz, Watamaniuk & McKee, 1995; Watamaniuk et al., 1995; also see Fig. 7 below). A large direction-selective detector would be poorly stimulated by circular trajectories, especially if it were large enough to see the trajectory change direction by 180°.

This failure of large detectors raises the question of whether several local motion units acting independently can explain the detectability of the signal trajectory. In order to test this hypothesis we measured the detectability of brief trajectories–those likely to stimulate local motion detectors. We then compared performance to the prediction of a local motion model. We also tested whether such a model could explain the detectability of extended motion trajectories. The experiments in this study are restricted to conditions that test specific hypotheses relating to the motion mechanisms underlying trajectory detection. A thorough investigation of the detectability of a trajectory in noise, as a function of noise dot density, size of displacement on each frame, and duration of presentation has already been done (Watamaniuk et al., 1995).

2. General methods

The stimuli were dots that subtended 2 arc min, at a viewing distance of 1 m. The signal dot moved in one of eight directions, randomly picked on each interval. On each frame it was displaced in a consistent direction by 0.17°. This corresponded to a velocity of 12°/s, given the 71 Hz frame rate of the display monitor. Each noise dot was displaced by the same amount as the signal on each frame, but its direction was randomly sampled from 360°. As the signal and noise dots had the same step size and frame rate, it was impossible to discriminate between them on the basis of a pair of frames.

The luminance of the background was 45 cd/m^2. We deliberately picked a high luminance value for the background to minimize the visibility of the decaying phosphor trace. The luminance of the dots was 130 cd/m^2 yielding a Michelson contrast of 50%. The display duration was either 100 or 200 ms, which corresponded to seven and 14 frames, respectively at the 71 Hz frame rate. These brief durations made it unlikely that observers had sufficient time to track the trajectory with smooth pursuit after detection of an initial segment. The display area was a circular region 12.6° in diameter. The number of dots in this area determined the dot density. In our experiments the total number of dots (signal + noise) was kept fixed at 380, which corresponded to a dot density of 3 dots/degree^2.

We used a 2-alternative forced choice procedure with two temporal intervals. One of the intervals, picked at random, contained the signal in noise, whereas the other interval contained the noise alone. Observers were asked to choose the interval with the signal. Feedback was provided.

2.1. Local motion model

Briefly, our local model consisted of a motion energy stage (Adelson & Bergen, 1985), followed by a decision stage. The motion energy stage was made up of units with a temporal integration time of 100 ms, tuned to one of eight preferred directions, and with a spatial scale optimally tuned for motion that occurred in 100 ms. The local stage was followed by a decision stage that implemented the maximum rule, i.e. it selected the largest output across directions and positions within both the signal and noise intervals, and then chose the interval with the larger value. We chose the non-opponent form of motion energy model because it is consistent with physiological evidence from complex cells in cat area 17 (Emerson, Bergen & Adelson, 1992), and because it has successfully been used to predict psychophysical motion thresholds (Anderson & Burr, 1989, 1991). Our choice of decision rule was based on psychophysical studies of uncertainty, where the maximum rule is better suited to detecting a signal in one of many locations than a decision rule that sums all the responses in an interval and chooses the interval with the larger value (Graham, Kramer & Yager, 1987; Verghese & Stone, 1995). Under conditions of uncertainty, as in the case of detecting a single signal in one of many locations, the maximum rule is nearly equivalent to the optimal model (Nolte & Jaarsma, 1967; Pelli, 1985). It is possible that some other class of motion model and decision rule would generate different predictions from those described here. However, this model and decision rule are the most common choices for modeling this type of psychophysical experiment.

2.2. The motion energy stage

The local motion energy unit was composed of two units with the same preferred direction, but whose spatio-temporal profile was 90° out of phase, i.e. a quadrature pair (Adelson & Bergen, 1985). The filters

were 3-D Gabors, i.e. Gaussians multiplied by a sinusoid in two spatial dimensions and one temporal dimension (Grzywacz & Yuille, 1990). The even (cosine) filter was of the form

$$\mathrm{Gab}_{\cos}(x, y, t) = \frac{1}{(2\pi)^{3/2}\sigma^2\sigma_t}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\exp\left(-\frac{t^2}{2\sigma_t^2}\right) \times \cos(w(x\cos\theta + y\sin\theta) + w_t t)$$

where σ and σ_t are the space and time constants of the spatial Gaussian (in x and y)[1] and the temporal Gaussian respectively, ω and ω_t are the spatial and temporal frequencies respectively, and θ is the direction of motion.

The odd (sine) filter was of the form;

$$\mathrm{Gab}_{\sin}(x, y, t) = \frac{1}{(2\pi)^{3/2}\sigma^2\sigma_t}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\exp\left(-\frac{t^2}{2\sigma_t^2}\right) \times \sin(w(x\cos\theta + y\sin\theta) + w_t t)$$

The input image was convolved with each of these filters, and the odd- and even-filter outputs were squared and summed to yield motion energy. The squaring operation made the filters insensitive to the phase of the signal. We assumed that each location had one such filter-pair tuned to one of eight directions, with their preferred directions 45° apart. The optimal spatial frequency of these filters was variable, while their peak temporal frequency was fixed at 8.31 Hz (Watson & Nachmias, 1977; Anderson & Burr, 1985; Hess & Snowden, 1992) The filters were assumed to have spatial and temporal bandwidths of 1.5 and four octaves respectively (Kulikowski & Tolhurst, 1973; Graham, 1989). The temporal filter had a time constant of 25.6 ms, which corresponded to an integration time of 100 ms (95% of the area under the filter). This choice of temporal parameter values was consistent with data from Watson (1979) and Festa and Welch (1997) as well as with data that we obtained in a concurrent temporal summation study for dot stimuli (Verghese, McKee & Grzywacz, 1997).

We simulated the response of the motion energy stage to our experimental stimuli for 10000 trials. Each trial consisted of two intervals, a signal + noise interval and a noise-only interval. For each trial we calculated a signal-to-noise ratio based on the response of a single unit positioned at the center of the signal dot's trajectory. We repeated this calculation for several values of the space constant of the underlying Gaussian. Fig. 1 plots the average signal-to-noise ratio over 10000 trials as a function of the space constant. The space constant is expressed in terms of the hop size–the displacement between successive frames. As the frame rate is fixed, increases in hop size correspond to increases in velocity. Thus, the representation in terms of multiples of hop size has the advantage of being independent of velocity. For the particular velocity used in our experiment, one hop corresponds to a space constant of 0.17°. The open circles in Fig. 1 represent the case when the local motion detector had an integration time of 100 ms, and the signal trajectory consisted of six hops (100 ms). This function has a broad peak in the vicinity of 2.1 hops, which we take as the space constant σ_{opt} of the optimum motion unit. In other words, the optimum unit has the best signal-to-noise ratio for the noise density used in this experiment. For comparison, we attempted to determine the σ_{opt} for a local detector with an integration time of 200 ms. The filled squares represent the case when the local detector had an integration time of 200 ms, and was stimulated with a trajectory that lasted 13 hops (200 ms). The signal-to-noise ratio in this case is much lower than in the case of a unit that integrates for 100 ms. Moreover, the function is flat and has no clear peak. Therefore it is not clear that there is an optimally-sized detector, with equal space constants for width and height, for a 200 ms trajectory.

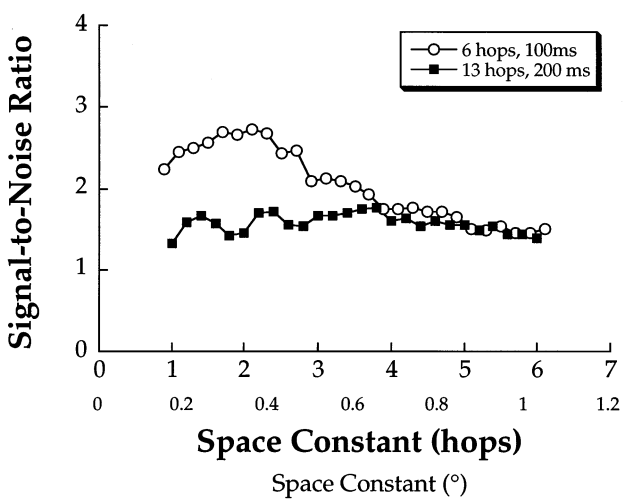

Fig. 1. The ratio of the response of a motion energy unit, centered on the trajectory, to a trial containing signal as well as noise to that of a trial containing noise alone, is averaged over 10000 trials and plotted as a function of the number of hops (displacement between successive frames). One hop corresponds to 0.17° for the velocity used in this experiment. The open circles plot the signal-to-noise ratio for a brief trajectory that lasted 100 ms (six hops), while the solid squares plot the signal-to-noise ratio for a trajectory that lasted 200 ms (13 hops).

[1] Estimates of the psychophysical receptive fields of motion sensitive units by both summation and masking studies indicate that the length-to-width ratio of these units is about 1 (Anderson & Burr, 1985, 1991). We therefore assumed that the local motion units had equal space constants along their length and width, i.e. in the directions parallel and orthogonal to the preferred orientation axis, respectively.

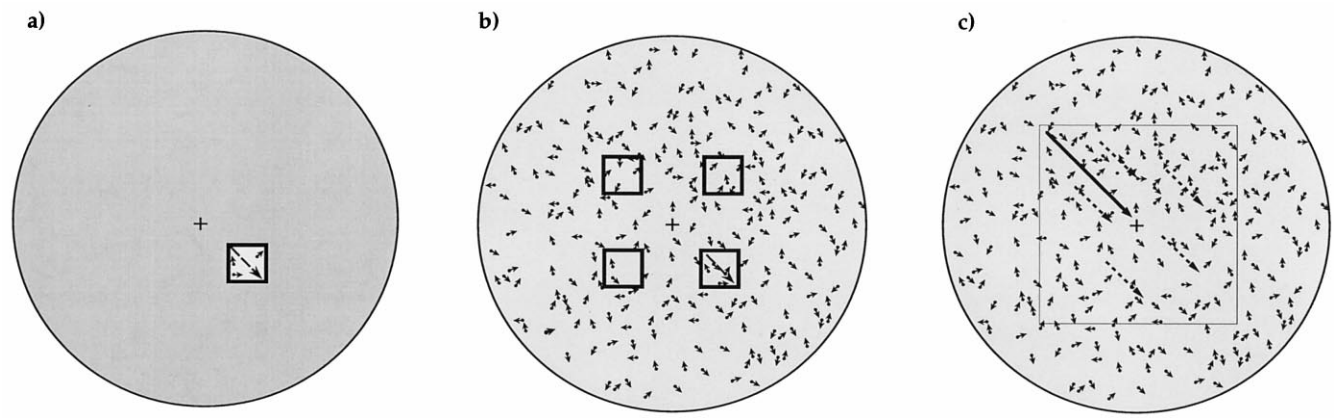

Fig. 2. Stimulus arrangements for detecting linear trajectories in noise. In all cases the large circle (not drawn to scale) indicates the stimulus display area. (a) Brief trajectory in a known location. The trajectory appeared in a square box 1° on the side, centered at 1.4° from fixation on the negative diagonal. The display area outside the box was masked off. The trajectory could move in one of eight directions. (b) Brief trajectory in an unknown location. The trajectory appeared in one of four boxes located on the diagonals and centered at 1.4° from fixation. Each box was a square 1° on the side. When the trajectory was constrained to appear in one location, it appeared in the lower right box, for two locations it appeared either in the upper left or lower right boxes, and for four locations it appeared in any one of the four boxes. Square outlines marked potential signal locations, but the noise in the entire display region was visible. (c) Comparison of brief and extended trajectories. The extended and brief trajectories, shown by the solid and dashed arrows, respectively, had durations of 200 and 100 ms respectively. Both were presented in a noise interval of 200 ms and were constrained to appear within a square window $4° \times 4°$, centered on fixation. (The square window is shown as an aid to visualize the display: it did not appear in the stimulus display.) The multiple 100 ms trajectories were presented either synchronously, or asynchronously within the 200 ms stimulus period.

For the current study we assumed that the area over which the stimulus could appear was tiled with optimally-sized motion detectors in a square lattice, with a uniform center-to-center spacing of 2σ. At each location there were detectors tuned to one of eight directions. The output of each motion-energy unit was assumed to be independent of other units. The motion-energy stage was followed by a decision stage that implemented the 'max' rule: in each interval, the largest response across detectors (at all locations and orientations) was selected and compared to the largest response in the other interval. The model was considered to have correctly detected the signal interval, if the larger of these two responses came from the signal interval.

3. Experiment 1: Brief signal in a known location

Since we were testing a local model, we started with a local version of the Watamaniuk et al. (1995) stimulus. To stimulate a local detector we constrained the stimulus to appear in a known location and restricted it to a duration of 100 ms. In this experiment we masked off the whole display except for the location in which the trajectory appeared. The mask had approximately the same luminance as the background. The window was $1 \times 1°$ and it was centered 1.4° from fixation, on the negative diagonal (see Fig. 2(a)). The trajectory passed through the center of the window in one of eight directions. Our choice of location was influenced by the fact that the trajectory stimulus in the original study (Watamaniuk et al., 1995) was generally presented at a parafoveal location. Observers were asked to maintain fixation and to judge which of two temporal intervals contained the signal. The experiment was run in blocks of 96 trials, and data are expressed as the average percent correct across two or four blocks of trials for observers SPM and PV, respectively. The error bars represent the standard error of the percent correct measure across blocks of trials.

3.1. Results

The percentage of correct responses for this condition are 82.3 ± 2.1 and $79.4 \pm 2.3\%$ for observers SPM and PV, respectively. As for the model, we assume that there is one optimally-sized detector, with an integration time of 100 ms, tuned to each of eight directions, centered on this location. The model predicts a performance of 82.8% correct, which is quite close to the data. The goodness of fit suggests that the added external noise was large compared to the internal noise of the human observer, given that the model unit had no internal noise. Thus it appears that the local motion energy unit can account for human detection of local motion signals in noise. Stronger evidence of the local model's ability to predict the detectability of local signals is presented in Experiments 2 and 3.

4. Experiment 2: Brief signal in multiple locations

In the original trajectory paradigm of Watamaniuk et al. (1995), the trajectory center was randomly located in a $2 \times 2°$ region centered on fixation. In order to approximate these conditions we added spatial uncertainty to the conditions of Experiment 1, and measured performance as a function of spatial uncertainty. Instead of a single location there were four possible locations in which the trajectory could appear. The squares in Fig. 2(b) represent those locations and were visible as outlines superimposed on the noise display. When the trajectory was constrained to appear in one location, it appeared in the lower right square, for two locations it appeared either in the upper left or lower right squares, and for four locations it appeared in any one of the four squares.

4.1. Results and discussion

Fig. 3 shows the data for this experiment. Proportion correct is plotted versus number of locations. Each point represents a minimum of 288 trials measured over three blocks of 96 trials each. The error bars represent the standard error of the percent correct measure across blocks of trials. The different symbols are data for three different observers. For all three observers, performance falls with increasing uncertainty about stimulus location. The dashed line is the model's prediction for this experiment. Once again we assumed that there was an optimal unit tuned to each of eight directions at the relevant locations. The model predicts a similar trend with increasing number of locations. This is the familiar effect of the increase in uncertainty, where increasing the number of noise locations increases the probability that the largest value in the noise-only interval exceeds the largest value in the signal + noise interval.

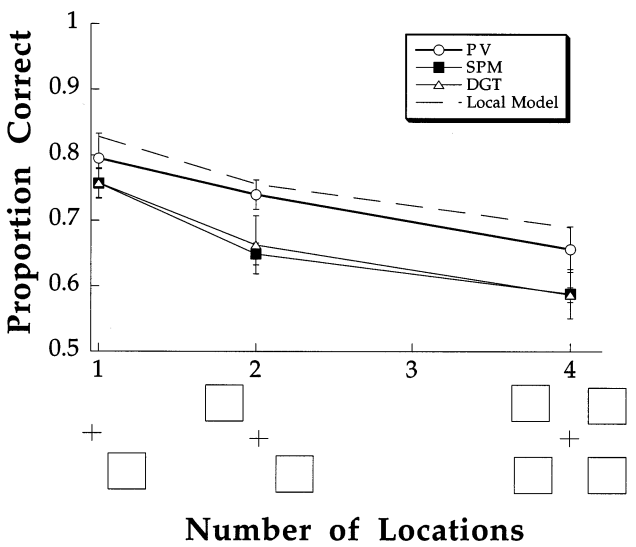

Fig. 3. Proportion correct as a function of the number of potential stimulus locations. The different symbols plot data for different observers. The dashed line plots the prediction of our local motion model as a function of potential signal locations.

The absolute performance of the model is better than the data. A comparison of the data for the single location case from Experiment 1 and this experiment shows that observers are better able to detect the stimulus trajectory in the former case, despite the fact that it appeared in the same location. This is probably due to the fact that observers were unable to completely ignore the noise outside the possible stimulus locations. In Experiment 1 the stimulus was restricted to $1° \times 1°$ window with the rest of the display masked off, while the entire display was visible in Experiment 2, with square outlines marking potential signal locations.

5. Experiment 3: A comparison of extended and local trajectories

We next considered more realistic trajectories–ones that excite more than one independent local unit. Given that our local model does a reasonable job of predicting human performance for detecting brief trajectories, can it predict the performance for longer trajectories? In other words, is a long trajectory simply the sum of its local parts? In the next experiment we compare detection of a single 200 ms trajectory to the detection of two or more 100 ms trajectories (Watamaniuk & McKee, 1993) to determine how many of these short trajectories are required to equal performance for the single long trajectory. In particular, we wanted to determine whether the detectability of the 200 ms trajectory could be predicted from two independent 100 ms trajectories.

Fig. 2(c) describes the stimulus that we used to compare the detectability of the long and short trajectories. Both the long trajectory (shown by the solid arrow) and short trajectories (shown by the dashed arrows) were constrained to fall within a square window 4°on each side, centered on fixation. (The square window was not visible in the experiment, and is shown in Fig. 2(c) only as an aid to visualize the display.) The long trajectory was 200 ms long. The trajectory pieces were 100 ms long, but occurred in a noise interval that was 200 ms long. With this design, both the 200 ms trajectory and the multiple 100 ms trajectories would have the same duration of noise. The trajectories could either appear synchronously or asynchronously. In the synchronous case they appeared simultaneously, at a random time within the 200 ms interval. In the asynchronous case they were staggered in time within the display interval. For instance in the case of two 100 ms trajectory pieces, they appeared in non-overlapping temporal intervals. This latter configuration more

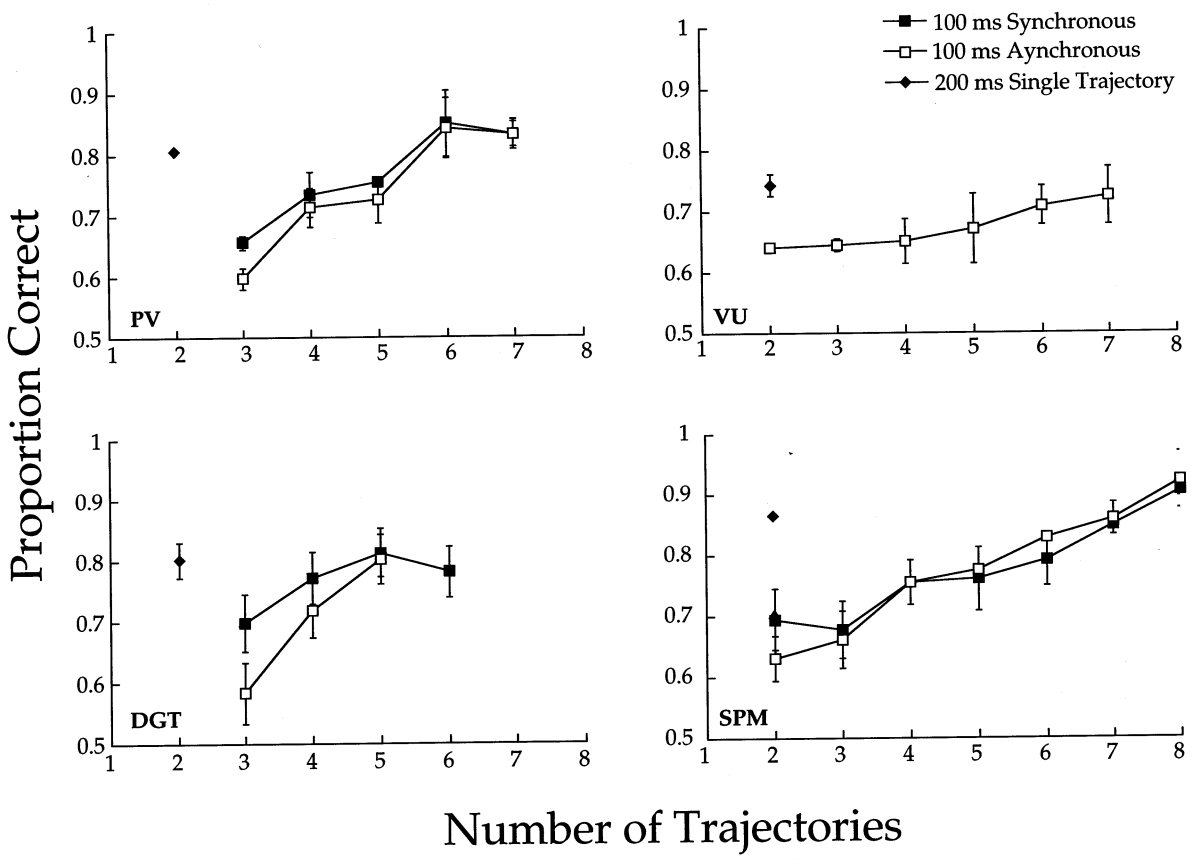

Fig. 4. Proportion correct for four observers as a function of the number of 100 ms trajectory pieces. The open and solid squares represent data for the multiple synchronous and asynchronous 100 ms trajectories, respectively. Proportion correct increases with the number of 100 ms trajectories. The solid diamond plots data for the single 200 ms trajectory. It takes about 5–7 100 ms trajectories to equal the detectability of a single 200 ms trajectory.

closely approximated the case of the single 200 ms trajectory.

5.1. Results and discussion

Fig. 4 plots data for this experiment. Proportion correct is plotted versus the number of 100 ms trajectories. The four panels show data for four observers. Each point represents the average of at least 192 trials measured over two blocks of trials. The error bars show the standard error of the percent correct estimate across blocks of trials. The open squares represent synchronous trajectories, and the solid squares represent asynchronous trajectories. Performance improves as the number of 100 ms trajectories increases. Observers PV and DGT have a higher probability of detection in the synchronous condition. As the locations of the trajectories were picked randomly, it is likely that in the synchronous condition, a subset of them appeared close to each other and were thus were more detectable (presumably because they appeared within the area of summation of a local detector).

The solid diamonds show the proportion correct for the 200 ms long trajectory. A comparison of the performance for the single 200 ms trajectory and two 100 ms trajectories indicates that the former is much more detectable. It takes five to seven 100 ms trajectories to equal performance with the longer trajectory.

Fig. 5 shows the results for this experiment averaged across four observers. The solid squares are data for the asynchronous condition, and the open diamond represents average data for the long trajectory. Proportion correct is plotted versus the number of trajectories. The stippled region indicates the range of performance for the single 200 ms trajectory (the width of the region represents ± 1 standard error). The human observer is much better at detecting the 200 ms trajectory (81% correct) than two non-overlapping 100 ms trajectories. Proportion correct for two 100 ms trajectories is around 63%[2] and reaches that of the single long trajectory only when there are about six 100 ms trajectories.

[2] The proportion correct for a single 100 ms trajectory in this experiment is much lower ($< 60\%$ correct) than in the earlier case of signal location known exactly (81%), as the stimulus in this experiment could appear anywhere within a four degree window centered on fixation.

Let us consider the response of the local model to the 100 ms trajectory. The dashed line in Fig. 5 is the prediction of the local model as a function of the number of 100 ms trajectories. The local detectors are optimally stimulated by the 100 ms trajectory pieces (Figs. 1 and 2). The model predictions demonstrate that the local model is indeed a good fit to the multiple presentations of 100 ms trajectories. Recall that the local model selects the largest response across all the detectors that tile the stimulus area, in both the signal + noise, and noise only presentations and selects the interval with the largest response. The performance of the model improves with the number of 100 ms trajectories, because the increased number of signals increases the probability that the largest response comes from the signal interval.

As for the 200 ms trajectory, it stimulates two non-overlapping detectors, given a temporal integration time of 100 ms. If there are more than two detectors, they would sample overlapping parts of the stimulus (signal + noise). As the noise in the stimulus is large compared to the noise associated with a motion unit (for evidence see Experiment 1), overlapping samples will have correlated inputs (and responses). Thus, we first consider the case when the 200 ms trajectory is detected by two non-overlapping, independent detectors, centered on the first and second halves of the trajectory, respectively. In that sense, the motion model with local detectors does not differentiate between a 200 ms trajectory and two non-overlapping 100 ms trajectories: for both cases it predicts about 63% correct. This is shown by the model prediction in Fig. 5.

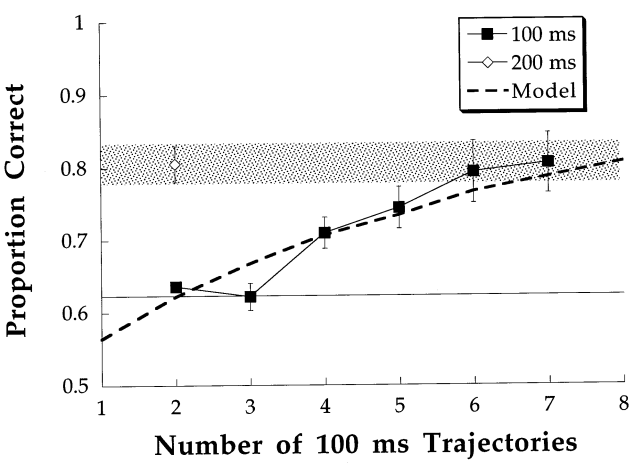

Fig. 5. Comparison of human and model performance. The solid squares and the open diamond plot the data, averaged across four observers, for the asynchronous 100 ms and the single 200 ms conditions, respectively. The stippled region indicates the standard error for the 200 ms condition. The dashed line plots the prediction of the local motion model as a function of the number of 100 ms trajectories. It predicts that the 200 ms trajectory is as detectable as two 100 ms trajectories. The model is a good fit to the 100 ms data, but severely underestimates the detectability of the 200 ms trajectory.

The model is a good fit to the data for two 100 ms trajectories, but its prediction for the 200 ms trajectory (shown by the solid horizontal line) falls well short of the observed performance in this task. A comparison of model performance in terms of d' will make this point more clearly. The model's d' for a single 100 ms trajectory is 0.3 (data not shown), and for two 100 ms trajectories is 0.45. This improvement is consistent with an optimal linear combination of two independent signals (Green & Swets, 1966). However, the observed proportion correct for a 200 ms trajectory corresponds to a d' of 1.24. These data indicate that trajectory detection cannot be explained by local independent motion detectors alone. Furthermore, the 3-fold improvement in detectability of a single 200 ms trajectory over two 100 ms trajectories, suggests that the pooling process that combines local-motion signals might incorporate a non-linear interaction between local units.

We also investigated the effect of tiling the 200 ms trajectory with overlapping detectors. This investigation was motivated by the fact that the receptive fields of cortical motion units overlap significantly. Moreover, as the number of overlapping units that sample the trajectory increases, the probability that one of them has fewer noise dots than average in its receptive field, also increases. We built a model with a total of seven overlapping motion energy units on the 200 ms trajectory, i.e. five units in between the ones sampling the first and second halves of the trajectory. This is the maximum number of motion detectors that would see different 100 ms segments (six displacements) of the 200 ms trajectory (13 displacements). Each unit integrated motion signals for 100 ms and its center was offset by one displacement step along the trajectory. We then measured the outputs of these seven overlapping units to 1000 trials, each consisting of a signal + noise interval and a noise-only interval. The units successfully detected the signal on a given trial if the largest response came from the signal interval. We repeated this procedure for the two non-overlapping units as well. A comparison of the results shows that on average, the percent correct for the overlapping-unit case is 70%, compared to 63% for the non-overlapping unit case. However, this an optimistic estimate of the advantage of overlap as we have only considered the effect of placing overlapping detectors on the trajectory, and not on the entire stimulus region. In general if there are n detectors covering the linear trajectory, then the number of detectors required to cover the square display region is proportional to n^2 Thus the proportion of detectors responding to noise increases with the degree of overlap, thus increasing the probability that the largest response comes from the noise interval. Furthermore, even the best case improvement from 63 to 70% with overlapping detectors is not sufficient to explain the improvement in detection from 63 to 81% in going

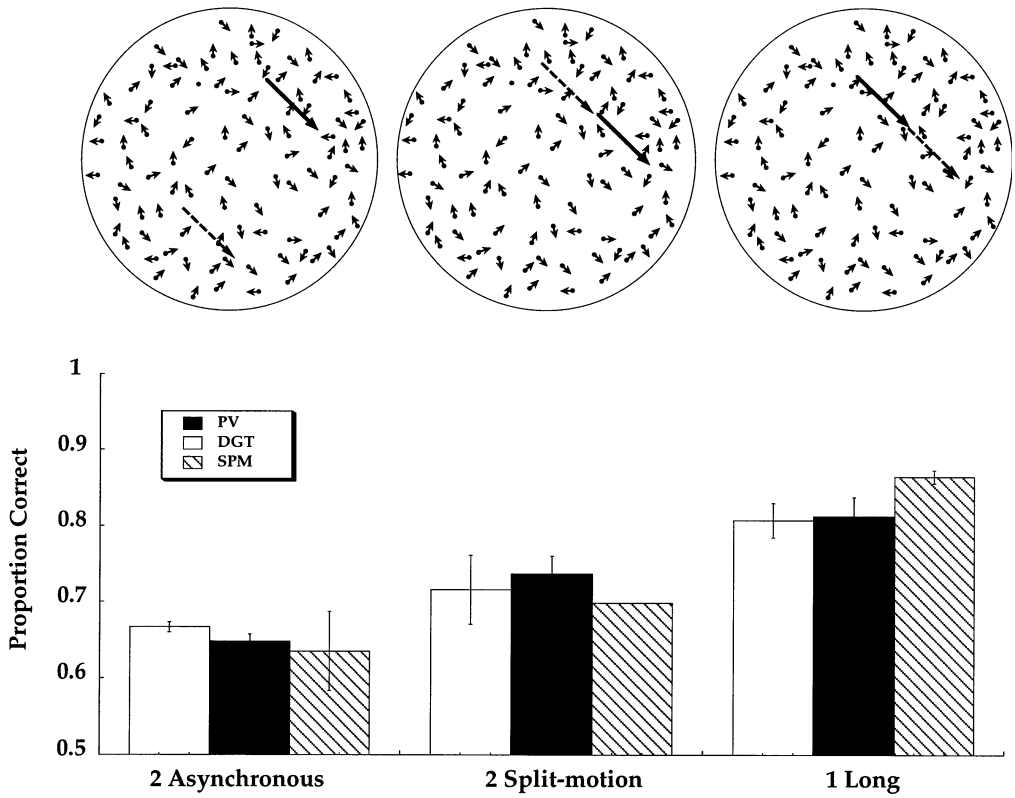

Fig. 6. Comparison of the detectability of two 100 ms trajectories as a function of their spatial and temporal relationship. The three configurations that were tested are shown above with the corresponding probabilities of detection plotted below. All three configurations were presented in 200 ms of noise. The solid and dashed arrows represent the trajectories that appeared in the first and second 100 ms periods, respectively. In the 'two asynchronous' case, the two 100 ms trajectories appeared at random locations within a square window $4° \times 4°$, centered on fixation. In the 'two split-motion' case, the two 100 ms pieces were in the same spatial configuration as an extended 200 ms trajectory, but they appeared in reversed temporal order. The 'one long' case had a single 200 ms trajectory.

from two asynchronous 100 ms trajectories to a single 200 ms trajectory. In conclusion, a model incorporating an important class of local units followed by the maximum rule cannot account for the improved detectability of long trajectories, even when detector overlap is assumed.

We also did a variation of the trajectory experiment to determine whether the detectability of the single 200 ms trajectory required a spatio-temporal sequence of signals or whether it was merely due to the spatial collinearity of the signals. To test the latter possibility, we reversed the temporal order of the two segments in the 200 ms trajectory, such that the second segment appeared first, followed by the first segment. The stimulus for this split-motion experiment is illustrated in Fig. 6, along with data for three observers. The first and second trajectory segments are represented by solid and dashed arrows, respectively. For comparison we have also plotted data for two asynchronous trajectories and for a single 200 ms trajectory. Data for each of these cases, was based on at least 288 trials per point, measured in three blocks of 96 trials each. The error bars represent the standard error of the percent correct measure over the blocks. Performance in the split-motion case is better than the asynchronous case but still much worse than the single 200 ms trajectory. The improvement over the asynchronous case suggests that observers benefited from knowing the location of the second 100 ms segment when they had information about the signal in the first 100 ms segment. An implementation of a local motion model that monitors all locations in the first part of the interval, finds the largest sample, and then monitors the location behind this location in the second part of the interval is consistent with the data. The fact that the split motion case is more detectable than two asynchronous trajectories shows that there is a benefit to knowing the location of the second 100 ms trajectory. However, this benefit is not sufficient to explain the detectability of a single 200 ms trajectory.

Thus an extended trajectory appears to be more detectable than the sum of its parts. Grzywacz et al. (1995) have suggested that the detection of such trajectories is subserved by a network that enhances motion signals in a consistent direction. This network has forward facilitation among units with similar direction preference, i.e. a unit facilitates another unit if its preferred direction points to that unit. Furthermore, as

these units have a direction tuning bandwidth of about 45°, they also facilitate the detection of a trajectory that changes direction slowly. This estimate of direction bandwidth is consistent with the results of Watamanuik et al. (1995), who measured the detectability of a trajectory that alternated direction by $\pm\theta$ around the direction of a linear trajectory, so that the trajectory changed direction by 2θ on each frame. The direction bandwidth (2θ) that corresponds to 75% correct was about 49°. Thus a circular trajectory that changes direction slowly would be good test of the flexible facilitation of the trajectory network model. In fact, such an experiment was performed by Watamanuik et al. (1995). These authors concluded that a circular trajectory that changed direction by 24° on every frame (and by 96° within the assumed temporal integration time of 100 ms) was as detectable as a linear trajectory centered at the same eccentricity. However this conclusion may be an artifact of their stimulus presentation. Their circular trajectory was concentric with fixation and therefore was always at the same eccentricity, while the linear trajectory was tangent to this circle, and therefore covered a larger range of eccentricities. In terms of our local model, a greater number of local units are required to tile the possible locations of the linear trajectory than in the case of the circular trajectory. Thus there was more uncertainty in the case of the linear trajectory which could have decreased its detectability relative to the circular trajectory.

6. Experiment 4: Circular trajectory

In this study we redid the experiment of Watamanuik et al. (1995) that measured the detectability of a 200 ms circular trajectory as a function of the radius of curvature. To make the number of locations that the observer had to monitor comparable to that for the case of the linear trajectory, the circular trajectory was not concentric with fixation. Specifically, the center of the circular arc forming the trajectory was constrained to appear randomly within ± 1°of fixation, and the trajectory could move in either a clockwise or a counter-clockwise direction[3]. Thus, the entire trajectory appeared in a region ± 2° from fixation. We measured the detectability of the trajectory as a function of its radius of curvature. The duration of the trajectory was 200 ms, brief enough so that the signal trajectory did not retrace its path even for the smallest radius of curvature used.

6.1. Results and discussion

The data are based on 192 trials per point for observers PV and DGT. Data for two observers (circles in Fig. 7) show that while trajectories with large radii of curvature were detected as well as linear trajectories (squares in Fig. 7), the trajectory with the smallest radius of curvature was harder to detect. Performance improves with increasing radius of curvature, asymptoting to that of a linear trajectory for radii larger than 1°. This indicates that observers can detect a trajectory that changes direction by 60° in 100 ms, or equivalently by 120° in 200 ms, almost as well as they can detect a linear trajectory. It would be hard to explain this detectability in terms of a single local detector, as it would require that the detector have an unusually large direction bandwidth, and an integration time that is longer than the 100–130 ms that has been reported (Watson, 1979; Fredericksen, Verstraten & van de Grind, 1994a; Watson & Turano, 1995). While the detectability of circular trajectories could be due to higher-level units that combine local signals consistent with rotary motion (as suggested by Sakata, Shibutani, Ito & Tsurugai, 1986; Tanaka & Saito, 1989; Anderson, Snowden, Treue & Graziano, 1990), such units cannot explain the detectability of trajectories that take a smooth, but wobbly motion path (Watamaniuk et al., 1995).

7. General discussion

Predictions based on the outputs of optimally tuned local motion energy detectors are consistent with the observed detectability of brief motion trajectories. This

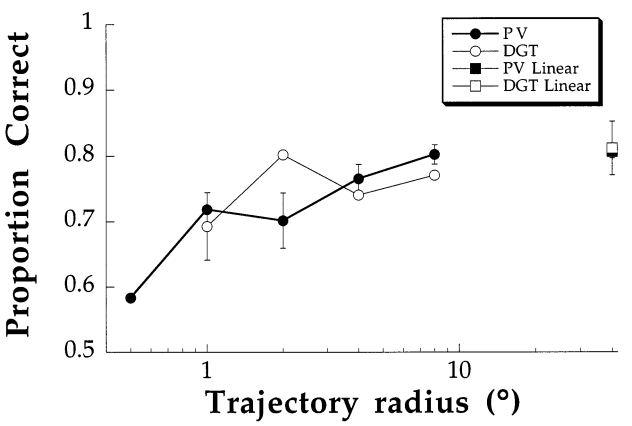

Fig. 7. Probability of detecting a circular trajectory as a function of its radius of curvature. The trajectory lay within a square window $4° \times 4°$, centered on fixation. The circles plot the circular trajectory data for two observers, while the squares plot comparable performance for a linear trajectory.

[3] In order to determine whether the trajectories for the different radii of curvature have comparable eccentricities from fixation, we measured the eccentricity of each position on a trajectory for the different radii of curvature. This was repeated for 1000 trials at each radius of curvature. Our analysis shows that the location of the trajectory points has the same distribution as a function of eccentricity for radii of curvature from 0.5 to 8°.

is true for brief trajectories in a known location, as well as unknown locations. It might appear that a larger motion detector that spans the entire trajectory might explain our data for longer trajectories. This is an unlikely explanation as noise constrains the scale of the local detector. Simulations with the local motion model show that for an integration time of 100 ms, the signal-to-noise ratio is an inverted U-shaped function of detector size, peaking for detectors with a space constant of 2.1 hops (see circles in Fig. 1). If however, one considers a detector large enough to see a 200 ms trajectory, it would also have to integrate noise signals for that long. The solid squares in Fig. 1 show that in this case, the signal-to-noise ratio is poor, and relatively flat across scale. If we take the peak of this function as four hops and tile the stimulus area with these larger (circular) motion units, then simulations show that the probability of detecting a 200 ms trajectory is even worse than if the stimulus area were tiled with motion units optimally tuned for 100 ms trajectories.

The success of optimally-sized local motion detectors at predicting the detectability of brief 100 ms trajectories suggests that such units can explain sequential recruitment effects that occur within the integration time of a motion detector. Recruitment refers to the finding that multi-frame motion is more detectable than two-frame motion spanning the same distance over time (Nakayama & Silverman, 1984; McKee & Welch, 1985). The improvement in detectability over longer durations (Snowden & Braddick, 1991) could very well be due to the interaction between such local mechanisms, as in the case of trajectory detection.

Although we have only considered trajectories of 100 and 200 ms duration, the data of Watamanuik et al. (1995) show that detectability continues to rise sharply for longer durations until it hits a ceiling. For instance, a 300 ms trajectory in comparable noise yields a percent correct of about 90%. This translates to a d' of about 1.81, while the predicted detectability for three independent 100 ms trajectories in a 300 ms display is 0.40.

While local, independent motion units do a good job of predicting the detectability of a short trajectory that falls within their receptive field, they appear to severely underestimate the detectability of an extended trajectory. This failure of local motion units implicates a process that nonlinearly combines their activity. Such a pooling of local signals could be achieved by a higher-level 'feature-tracking' system proposed by Lu and Sperling (1995) that tracks the motion of salient features. Alternatively, the activity of local units could be combined in a trajectory detection network that enhances motion signals in a consistent direction (Grzywacz, Smith & Yuille, 1989; Hubbard & Marshall, 1994; Grzywacz et al., 1995). In Grzywacz et al. (1995) a motion unit feeds forward facilitation to another motion unit if its preferred direction roughly points to that unit, and if that unit has a similar preferred direction. Thus a unit responding to a signal moving along a trajectory not only responds to the local motion of the signal, but receives feedforward activation from a similar unit at a preceding location along the trajectory. The model by Yuille, Burgi and Grzywacz (1998) uses a Bayesian probabilistic approach, where the enhanced response to a trajectory is mediated by an increased probability of a given speed and direction at a given location along a trajectory.

It can be argued that such networks are not necessary and that a detector that is spatially elongated along the direction of motion will be better tuned to motion along a trajectory. This is because such a detector integrates most of the signal and a relatively small proportion of the noise as compared to a detector with a circular profile whose spatial extent is large enough to cover the signal. However, there are three lines of reasoning that argue against elongated detectors. Firstly, psychophysical evidence from both summation and masking studies indicate that the psychophysical receptive fields of motion sensitive units have a length to width ratio of one (Anderson & Burr, 1985; Anderson, Burr & Morrone, 1991). These authors also estimate that the length-to-width ratio of motion-insensitive units is about 1.8, which is consistent with estimates from physiology for the length-to-width ratio of simple cells in cat striate cortex (Daugman, 1985; Webster & De Valois, 1985; Jones & Palmer, 1987) and is a validation of their psychophysical technique for measuring aspect ratios. There are several studies using the paradigm of van Doorn and Koenderink (1984) that suggest that receptive fields may be elongated in the direction of motion (Fredericksen, Verstraten & van de Grind, 1994b; van de Grind, Koenderink & van Doorn, 1986) . These studies measured detection thresholds for a rectangular random dot patch moving coherently in dynamic white noise, and showed that thresholds were lower when the patch was elongated in the direction of motion rather than when it was elongated in the direction perpendicular to motion. In fact, the more detailed studies of Fredericksen et al. (1994b) show that as the width of the patch is increased, thresholds improve at a faster rate than they improve when the height of the patch is increased. If one assumes that the improvement in the increasing-height condition is due to linear summation within a unit, then the improvement in the increasing-width condition is an accelerating non-linearity. This is similar to our data that shows nonlinear facilitation for extended trajectories: the 200 ms trajectory is at least a factor of three times more detectable than a 100 ms trajectory. Thus while the argument for motion units elongated in the direction of motion is plausible, one still has to account for the non-linear facilitation that is

supposedly occurring within a unit.

A second line of reasoning against elongated detectors is the following: if they are the explanation for the detectability of extended trajectories, then they should also have temporal integration times that span the trajectory duration. However, psychophysical estimates of temporal integration times range from 100 to 133 ms (Watson, 1979; Fredericksen et al., 1994b; Watson & Turano, 1995), which are shorter than the 200 ms duration of the extended trajectory in our experiments. Furthermore, estimates of the temporal integration times of direction-selective cells in areas V1 and MT of macaques are of the order of 114 and 91 ms, respectively (Mikami, Newsome & Wurtz, 1986). The lack of evidence for long integration times argues against the detectability of extended trajectories being mediated by units that are spatially elongated in the direction of motion.

The third argument against elongated detectors is based on earlier results by Watamanuik et al. (1995) as well as our experiments on the circular trajectory indicating that a trajectory that changes direction gradually is detected as well as a linear trajectory. For instance, a circular trajectory that changes direction by 60° in 100 ms is easily detected. To postulate a pool of detectors tuned to circular trajectories of varying curvature, in addition to detectors tuned to linear trajectories could lead to a combinatorial explosion of the number of motion detectors at each location. Instead a more parsimonious explanation is that the detection of extended trajectories is mediated by a flexible network of detectors that propagate activation among units tuned to similar directions of motion. In fact, we suggest that the trajectory network builds the ideal detector for an extended trajectory. It creates an effective motion unit that is that is elongated in space-time, yet flexible enough to accommodate smooth changes in direction and speed.

8. Conclusions

We have presented further evidence for the enhanced detectability of extended trajectories. Moreover, we have shown that this detectability cannot be predicted by an important class of independent local motion units followed by the maximum decision rule. These results point to the existence of a specialized network that combines motion signals in a consistent direction, which presumably evolved to track objects moving along smoothly changing motion paths in the midst of conflicting motion signals.

Acknowledgements

A portion of these results were presented at the annual meeting of the Optical Society of America, 1996. We thank Doug Taylor for programming the experiments, Ol Braddick for suggesting the split-motion experiment, and Brent Beutter and Jack Nachmias for their comments on an earlier version of the manuscript. This research was supported by AFOSR Grant F49620-95-1-0265 and NEI Core Grant EY06883. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the US Government.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, *2*, 284–299.

Andersen, R. A., Snowden, R. J., Treue, S., & Graziano, M. (1990). Hierarchical processing of motion in the visual cortex of monkey. *Cold Spring Harbor Symposium on Quantitative Biology*, *LV*. Cold Spring Harbor Laboratory Press.

Anderson, S. J., & Burr, D. C. (1985). Spatial and temporal selectivity of the human motion detection system. *Vision Research*, *25*, 1147–1154.

Anderson, S. J., & Burr, D. C. (1989). Receptive field properties of human motion defector units inferred from spatial freguency masking. *Vision Research*, *29*, 1343–1358.

Anderson, S. J., & Burr, D. C. (1991). Spatial summation properties of directionally selective mechanisms in human vision. *Journal of the Optical Society of America A*, *8*, 1330–1339.

Anderson, S. J., Burr, D. C., & Morrone, M. C. (1991). Two-dimensional spatial and spatial-frequency selectivity of motion-sensitive mechanisms in human vision. *Journal of the Optical Society of America A*, *8*, 1340–1351.

Anstis, S. M., & Ramachandran, V. S. (1987). Visual inertia in apparent motion. *Vision Research*, *27*, 755–764.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, *2*, 1160–1169.

van Doorn, A. J., & Koenderink, J. J. (1984). Spatiotemporal integration in the detection of coherent motion. *Vision Research*, *24*, 47–53.

Emerson, R. C., Bergen, J. R., & Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research*, *32*, 203–218.

Festa, E. K., & Welch, L. (1997). Recruitment mechanisms in speed and fine-direction discrimination tasks. *Vision Research*, *37*, 3129–3143.

Fredericksen, R. E., Verstraten, F. A. J., & van de Grind, W. A. (1994a). An analysis of the temporal integration mechanism in human motion perception. *Vision Research*, *34*, 3153–3170.

Fredericksen, R. E., Verstraten, F. A. J., & van de Grind, W. A. (1994b). Spatial summation and its interaction with the temporal integration mechanism in human motion perception. *Vision Research*, *34*, 3171–3188.

Graham, N. V. S. (1989). *Visual pattern analyzers press*. New York: Oxford University.

Graham, N., Kramer, P., & Yager, D. (1987). Signal detection models for multidimensional stimuli: probability distribution and combination rules. *Journal of Mathematical Psychology*, *31*, 366–409.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

van de Grind, W. A., Koenderink, J. J., & van Doorn, A. J. (1986). The distribution of human motion detector properties in the monocular visual field. *Vision Research*, *26*, 797–810.

Grzywacz, N.M., Smith, J.A., & Yuille, A.L. (1989). A common theoretical framework for visual motion's spatial and temporal coherence. The Proceedings of the IEEE Workshop on Visual Motion (pp. 148–155). Washington, IEEE Computer Society.

Grzywacz, N. M., Watamaniuk, S. N. J., & McKee, S. P. (1995). Temporal coherence theory for the detection and measurement of visual motion. *Vision Research*, *35*, 3183–3203.

Grzywacz, N. M., & Yuille, A. L. (1990). A model for the estimate of local image velocity by cells in the visual cortex. *Proceedings of the Royal Society of London*, *B*, *239*, 129–161.

Hess, R. F., & Snowden, R. J. (1992). Temporal properties of human visual filters: number, shapes, and spatial covariation. *Vision Research*, *32*, 47–59.

Hubbard, R.S., & Marshall, J.A. (1994). Self-organizing neural network model of the visual inertia phenomenon in motion perception (Technical Report No. 94-001). University of North Carolina, Department of Computer Science, Chapel Hill, p. 26.

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, *58*, 1233–1258.

Kulikowski, J. J., & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *Journal of Physiology*, *232*, 149–162.

Lu, Z-L., & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research*, *35*, 2697–2722.

McKee, S. P., & Welch, L. (1985). Sequential recruitment in the discrimination of velocity. *Journal of the Optical Society of America A*, *2*, 243–251.

Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. *Journal of Neurophysiology*, *55*, 1328–1351.

Nakayama, K., & Silverman, G. H. (1984). Temporal and spatial characteristics of the upper displacement limit for motion in random dots. *Vision Research*, *24*, 293–299.

Nolte, L. W., & Jaarsma, D. (1967). More on the detection of one of M orthogonal signals. *Journal of the Acoustical Society of America*, *41*, 497–505.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, *2*, 1508–1532.

Ramachandran, V. S., & Anstis, S. M. (1983). Extrapolation of motion path in human visual perception. *Vision Research*, *23*, 83–85.

Sakata, H., Shibutani, H., Ito, Y., & Tsurugai, K. (1986). Parietal cortical neurons responding to rotary movement of visual stimulus in space. *Experimental Brain Research*, *61*, 658–663.

Snowden, R. J., & Braddick, O. J. (1989). The combination of motion signals over time. *Vision Research*, *29*, 1621–1630.

Snowden, R. J., & Braddick, O. J. (1991). The temporal integration and resolution of velocity signals. *Vision Research*, *31*, 907–914.

Tanaka, K., & Saito, H. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology*, *62*, 626–641.

Verghese, P., & Stone, L. S. (1995). Combining speed information across space. *Vision Research*, *35*, 2811–2823.

Verghese, P., McKee, S. P., & Grzywacz, N. M. (1997). Stimulus configuration determines the detectability of motion signals in noise. *Investigative Ophthalmology & Visual Science*, *38*, 215.

Watamaniuk, S. N. J., & McKee, S. P. (1993). Why is a trajectory more detectable in noise than correlated signal dots? *Investigative Ophthalmology & Visual Science*, *34*, 1364.

Watamaniuk, S. N. J., McKee, S. P., & Grzywacz, N. M. (1995). Detecting a trajectory embedded in random-direction motion noise. *Vision Research*, *35*, 65–77.

Watson, A. B. (1979). Probability summation over time. *Vision Research*, *19*, 515–522.

Watson, A. B., & Nachmias, J. (1977). Patterns of temporal interaction in the detection of gratings. *Vision Research*, *17*, 893–902.

Watson, A. B., & Turano, K. (1995). The optimal motion stimulus. *Vision Research*, *35*, 325–336.

Webster, M. A., & De Valois, R. L. (1985). Relationship between spatial-frequency and orientation tuning of striate cortical cells. *Journal of the Optical Society of America A*, *2*, 1124–1132.

Yuille, A. L., Burgi, P-Y., & Grzywacz, N. M. (1998). Visual motion estimation and prediction: a probabilistic network model for temporal coherence. In T. Watanabe, *High-level motion processing*: *from computational*, *neurophysiological and psychophysical perspectives*. Cambridge: MIT Press.